

Swiss Institute of
Bioinformatics

Advanced R September 2024

Frédéric Schütz (Frederic.Schutz@sib.swiss)

!



?

Are these valid R commands ?

If yes, what do they do ?

If no, why ?

c=c (c=c)

c=c (c="c")

BIostatISTICS PLATFORM

Faculty of Biology and Medicine, UNIL

[Home](#)[Presentation](#)[Services](#)[Collaborations](#)[Team and contact](#)[More information ▾](#)

<https://wp.unil.ch/biostatistics/>

HOME

The Biostatistics platform of the Faculty of Biology and Medicine acts as a main entry point for all your questions related to biostatistics and data analysis. If needed, the platform can redirect you to the people with the best competences in order to answer your questions.

Contact: Frederic.Schutz@unil.ch



Schedule

9:00 - 10:30

10:30 - 10:45 break

10:45 - 12:15

12:15 - 13:30 break

13:30 - 15:00

15:00 - 15:15 break

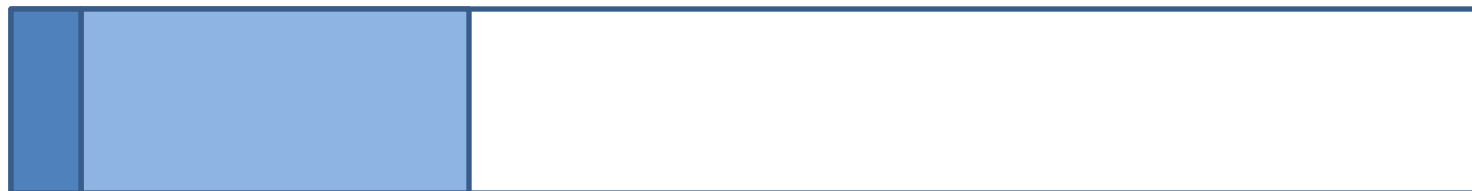
15:15 - 16:45

17:00 end of day

Introduction
Reproducible
research



Introduction
Reproducible
research



Reminder about R
data structures

Introduction
Reproducible
research

Efficient programming



Reminder about R
data structures

Introduction
Reproducible
research

Efficient programming



Reminder about R
data structures

Data manipulation

Content (first part)

Introduction
Reproducible
research

Efficient programming

R best practices



Reminder about R
data structures

Data manipulation

- Hadley Wickham. «Advanced R». CRC Press, 2014
- Phil Spector. «Data manipulation with R». Springer, 2008
- W. Venables and B. Ripley. «S Programming». Springer, 2004
- John M. Chambers. «Software for Data Analysis – Programming with R». Springer, 2008

Base R

vs

Tidyverse

I



?

Reproducible research

**The
Economist**

OCTOBER 19TH-20TH 2013

economist.com

Washington's lawyer surplus

How to do a nuclear deal with Iran

Investment tips from Nobel economists

Junk bonds are back

The meaning of Sachin Tendulkar

HOW SCIENCE GOES WRONG

The Economist, October 2013

THIS WEEK

EDITORIALS

FOOD Fertilizer is the best way to feed Africa right now **p.510**



WORLD VIEW Teach young scientists how to manage their labs **p.511**

VENICE City of Water edges closer to becoming a city under water **p.512**

Must try harder

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

Science: Branch of knowledge or study dealing with a body of facts or truths systematically arranged. So says the dictionary. But, as most scientists appreciate, the fruits of what is called science are occasionally anything but. Most of the time, when attention focuses on divergence from this gold (and linguistic) standard of science, it is fraud and fabrication — the facts and truth — that are in the spotlight. These remain important problems, but this week *Nature* highlights another, more endemic, failure — the increasing number of cases in which, although the facts and truth have been established, scientists fail to make sure that they are systematically arranged. Put simply, there are too many careless mistakes creeping into scientific papers — in our pages and elsewhere.

A Comment article on page 531 exposes one possible impact of such carelessness. Glenn Begley and Lee Ellis analyse the low number of cancer-research studies that have been converted into clinical success, and conclude that a major factor is the overall poor quality of published pre-clinical data. A warning sign, they say, should be the “shocking” number of research papers in the field for which the main findings could not be reproduced. To be clear, this is not fraud — and there can be legitimate technical reasons why basic research findings do not stand up in clinical work. But the overall impression the article leaves is of insufficient thoroughness in the way that too many researchers present their data.

The finding resonates with a growing sense of unease among specialist editors on this journal, and not just in the field of oncology. Across the life sciences, handling corrections that have arisen from avoidable errors in manuscripts has become an uncomfortable part of the publishing process.

The evidence is largely anecdotal. So here are the anecdotes: unrelated data panels; missing references; incorrect controls; undeclared cosmetic adjustments to figures; duplications; reserve figures and dummy text included; inaccurate and incomplete methods; and improper use of statistics — the failure to understand the difference between technical replicates and independent experiments, for example.

It is usually the case that original data can be produced, mistakes corrected, and the findings of the corrected research paper still stand. At the very least, however, there is too little attention paid and too many corrections, which reflect unacceptable shoddiness in laboratories that risks damaging trust in the science that they, and others, produce.

The situation throws up many questions. Here are three of them. Who is responsible? Why is it happening? How can it be stopped?

The principal investigators (PIs) of any lab from which the work originates, especially if their names are on the paper, have an absolute and unavoidable responsibility to ensure the quality of the data from their labs, even if the main work is done by experienced postdocs. Officially, postdocs and graduate students are still in training, and it is the PI's job to make sure they are properly trained — in statistics and appropriate image editing, for a start. It is unacceptable for lab heads — who are happy to take the credit for good work — to look at raw data

for the first time only when problems in published studies are reported.

In private, scientists who run labs in even the most prestigious universities admit that they have little time to supervise and train all their students. Institutions such as the European Molecular Biology Laboratory in Heidelberg, Germany, have maximum lab sizes for this reason. Funding agencies should require grant applicants to indicate lab size and offer adequate supervision. As is the case in commercial companies,

larger labs should introduce formal training and a management hierarchy, with more experienced postdocs and research associates required to sign off data and experiments if PIs cannot do so themselves.

“Handling corrections that have arisen from avoidable errors in manuscripts has become an uncomfortable part of the publishing process.”

What can journal editors and referees do? Sloppiness is sometimes caught, but so much must be taken on trust. Journals should certainly offer online commenting, so that alert readers can point out errors.

Where comments or corrections appear in other journals, these should be linked from the original paper — as the Comment authors recommend.

There should also be increased scope to publish fuller results from an experiment, and subsequent negative or positive corroborations. There is an opportunity here for ‘minimum threshold’ journals, such as *PLoS ONE* and *Scientific Reports*. Editors and referees cannot be expected to divine when only positive data are included and inconvenient results left out, but journals should encourage online presentation of the complete picture. And scientists should offer it. The complete picture is, after all, what this science of ours strives to provide. ■

THIS WEEK

EDITORIALS

FOOD Fertilizer is the best way to feed Africa right now **p.510**



WORLD VIEW Teach young scientists how to manage their labs **p.511**

VENICE City of Water edges closer to becoming a city under water **p.512**

Must try harder

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

Science: Branch of knowledge or study dealing with a body of facts or truths systematically arranged. So says the dictionary. But, as most scientists appreciate, the fruits of what is called science are occasionally anything but. Most of the time, when attention focuses

for the first time only when problems in published studies are reported.

In private, scientists who run labs in even the most prestigious universities admit that they have little time to supervise and train all their students. Institutions such as the European Molecular Biology Labora-

Must try harder

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

...institutions to the way the laboratory researchers present their data.

The finding resonates with a growing sense of unease among specialist editors on this journal, and not just in the field of oncology. Across the life sciences, handling corrections that have arisen from avoidable errors in manuscripts has become an uncomfortable part of the publishing process.

The evidence is largely anecdotal. So here are the anecdotes: unrelated data panels; missing references; incorrect controls; undeclared cosmetic adjustments to figures; duplications; reserve figures and dummy text included; inaccurate and incomplete methods; and improper use of statistics — the failure to understand the difference between technical replicates and independent experiments, for example.

It is usually the case that original data can be produced, mistakes corrected, and the findings of the corrected research paper still stand. At the very least, however, there is too little attention paid and too many corrections, which reflect unacceptable shoddiness in laboratories that risks damaging trust in the science that they, and others, produce.

The situation throws up many questions. Here are three of them. Who is responsible? Why is it happening? How can it be stopped?

The principal investigators (PIs) of any lab from which the work originates, especially if their names are on the paper, have an absolute and unavoidable responsibility to ensure the quality of the data from their labs, even if the main work is done by experienced postdocs. Officially, postdocs and graduate students are still in training, and it is the PI's job to make sure they are properly trained — in statistics and appropriate image editing, for a start. It is unacceptable for lab heads — who are happy to take the credit for good work — to look at raw data

...an experiment, and subsequent negative or positive conclusions.

There is an opportunity here for 'minimum threshold' journals, such as *PLoS ONE* and *Scientific Reports*. Editors and referees cannot be expected to divine when only positive data are included and inconvenient results left out, but journals should encourage online presentation of the complete picture. And scientists should offer it. The complete picture is, after all, what this science of ours strives to provide. ■

Nature editorial: "Must try harder"

Glenn Begley and Lee Ellis analyze the low number of cancer-research studies that have been converted into clinical success, and conclude that **a major factor is the overall poor quality of published preclinical data.**

[...]

The overall impression the article leaves is of ***insufficient thoroughness in the way that too many researchers present their data.***

Reproducible Research:

“Research is reproducible if it can be reproduced by others”

Reproducible research

Of course, rerunning an experiment will give different results—an observation that gave rise to the development of statistics as a discipline.

Our focus here is “reproducible research” (RR) in the sense of reproducing conclusions from a single experiment based on the measurements from that experiment.



A **complete description** of the data and the analysis of that data — including computer programs — so the **results can be exactly reproduced** by others.



Reproducible
Research

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

BY KEITH A. BAGGERLY^{*} AND KEVIN R. COOMBES[†]

U.T. M.D. Anderson Cancer Center

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results, leading to exercises in “forensic bioinformatics” where aspects of raw data and reported results are used to infer what methods must have been employed. Unfortunately, poor documentation can shift from an inconvenience to an active danger when it obscures not just methods but errors. In this report, we examine several related papers purporting to use microarray-based signatures of drug sensitivity derived from cell lines to predict patient response. Patients in clinical trials are currently being allocated to treatment arms on the basis of these results. However, we show in five case studies that the results incorporate several simple errors that may be putting patients at risk. One theme that emerges is that the most common errors are simple (e.g., row or column offsets); conversely, it is our experience that the most simple errors are common. We then discuss steps we are taking to avoid such errors in our own investigations.

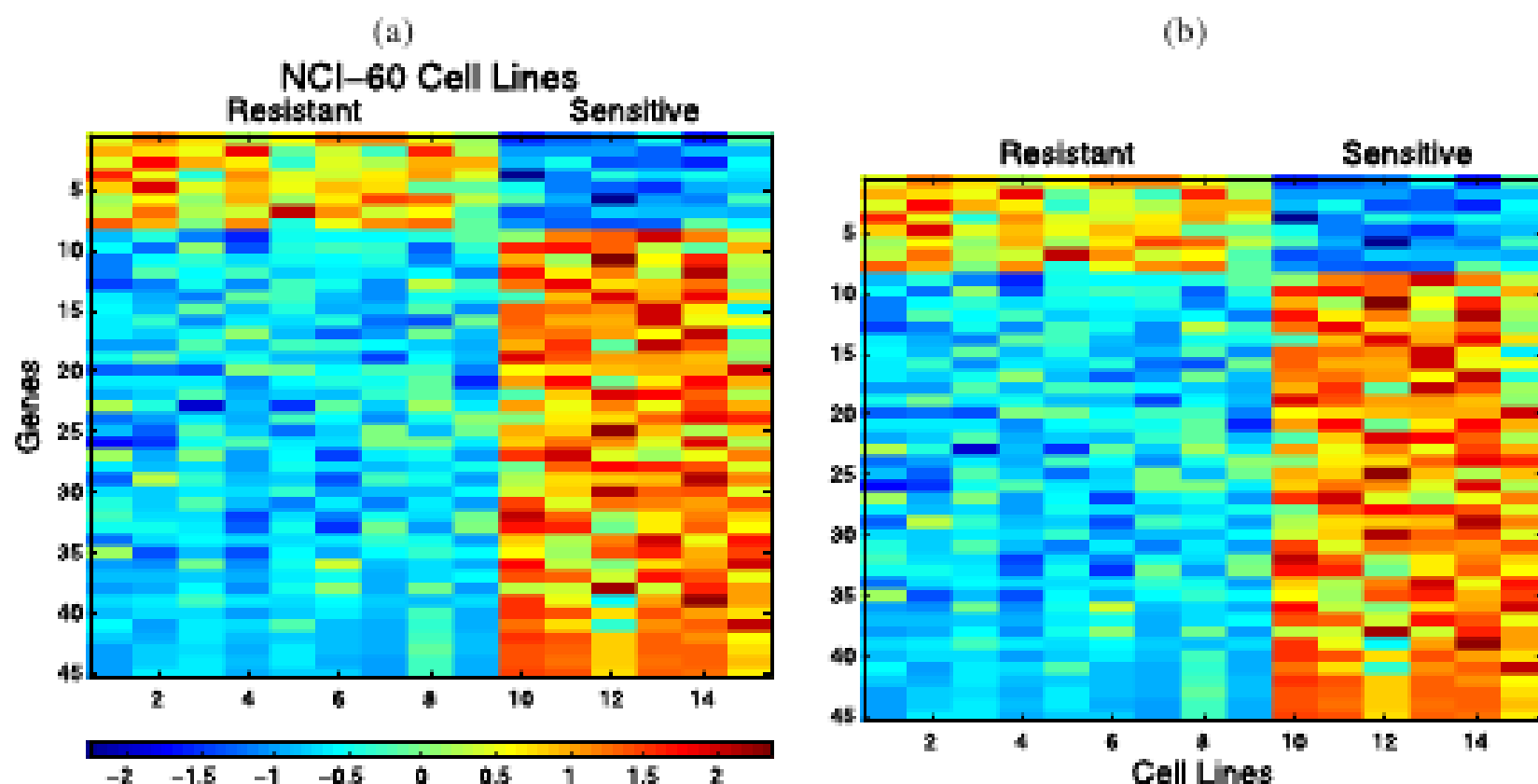


FIG. 4. Approximations to (a) the heatmap initially presented in Figure 4A of Augustine *et al.* (2009) for temozolomide, with lines reportedly chosen from the NCI-60 cell line panel, and (b) the heatmap presented in Figure 1 of Hsu *et al.* (2007) for cisplatin, with cell lines chosen from the 30-line panel of Györfy *et al.* (2006). The heatmaps are the same. We have independently generated the cisplatin heatmap using the Györfy *et al.* (2006) data, but the temozolomide heatmap is neither for temozolomide nor from the NCI-60 panel.

Conclusions from Baggerly and Coombes

"Poor documentation led a **report on drug A** to include a **heatmap for drug B** and a **gene list for drug C**.

These results are based on simple visual inspection and counting, and are not documented further."

Corrections made in the journal led to **further problems**.

Conclusions of Baggerly and Coombes

- The most common problems are simple:
 - confounding in the experimental design
 - mixing up the gene labels (off-by-one errors)
 - mixing up the group labels (sensitive/resistant)
- Most of these mix-ups involve simple switches or offsets.
- These mistakes are easy to make, particularly if working with Excel
- ... and/or if working with 0/1 labels instead of names

**We know we will make mistakes.
So...**

1) We should work in a way that will reduce the number of errors



**NO SMOKING
IN LAVATORY**

C1160-03-187



**NO SMOKING
IN LAVATORY**

C1160-03-187

**NO SMOKING
IN LAVATORY**

C1160-03-197



C1165-01-196

"Regardless of whether smoking is allowed in any other part of the airplane, lavatories must have self-contained, removable ashtrays located conspicuously on or near the entry side of each lavatory door "

US Code of Federal Regulations for airworthiness

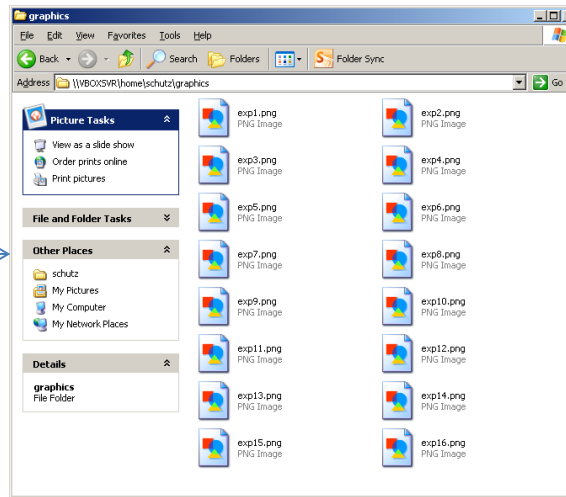
<https://www.law.cornell.edu/cfr/text/14/25.853>



**We know we will still
make mistakes !**

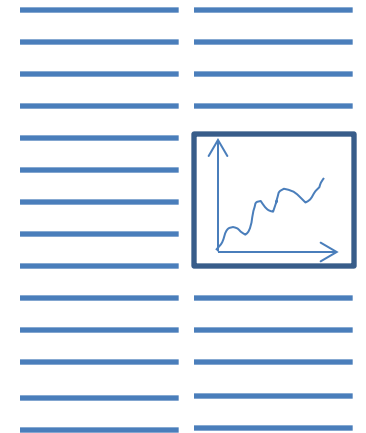
2) We must be able to track what we have done, so we can later find mistakes and correct them

Graphics



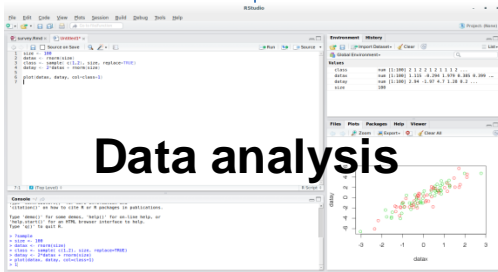
Manually

Article/Report



Manually

Data analysis



Stats results

```
H2-D -- CD8
              Df Sum Sq Mean Sq F value    Pr(>F)
Nlr5          1   33571   33571 343.424 1.71e-14 ***
CIITA          1     77     77  0.785  0.386
Nlr5:CIITA     1    609    609  6.225  0.021 *
Residuals    21   2053     98
---
H2-D -- CD4
              Df Sum Sq Mean Sq F value    Pr(>F)
Nlr5          1   31499   31499 90.402 4.65e-09 ***
CIITA          1    3467    3467  9.951  0.00478 **
Nlr5:CIITA     1     447     447  1.283  0.27009
Residuals    21   7317    348
---
H2-D -- NK
              Df Sum Sq Mean Sq F value    Pr(>F)
Nlr5          1  18867   18867 70.462 3.78e-08 ***
CIITA          1    386     386  1.441  0.243
Nlr5:CIITA     1    403     403  1.506  0.233
Residuals    21   5623    268
```

**First steps:
keep track of everything you do,
and all results you get**

Save your commands in a script.

Ideally, use Subversion or GIT to manage the revisions of this file

When you run it, save everything it produces (or at least the important results) to a text file.

At the beginning of the session, use `sessionInfo()` to record the software used.

"FINAL".doc



FINAL.doc!



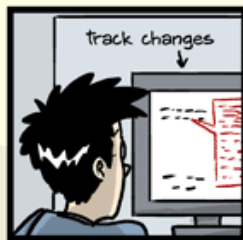
FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc

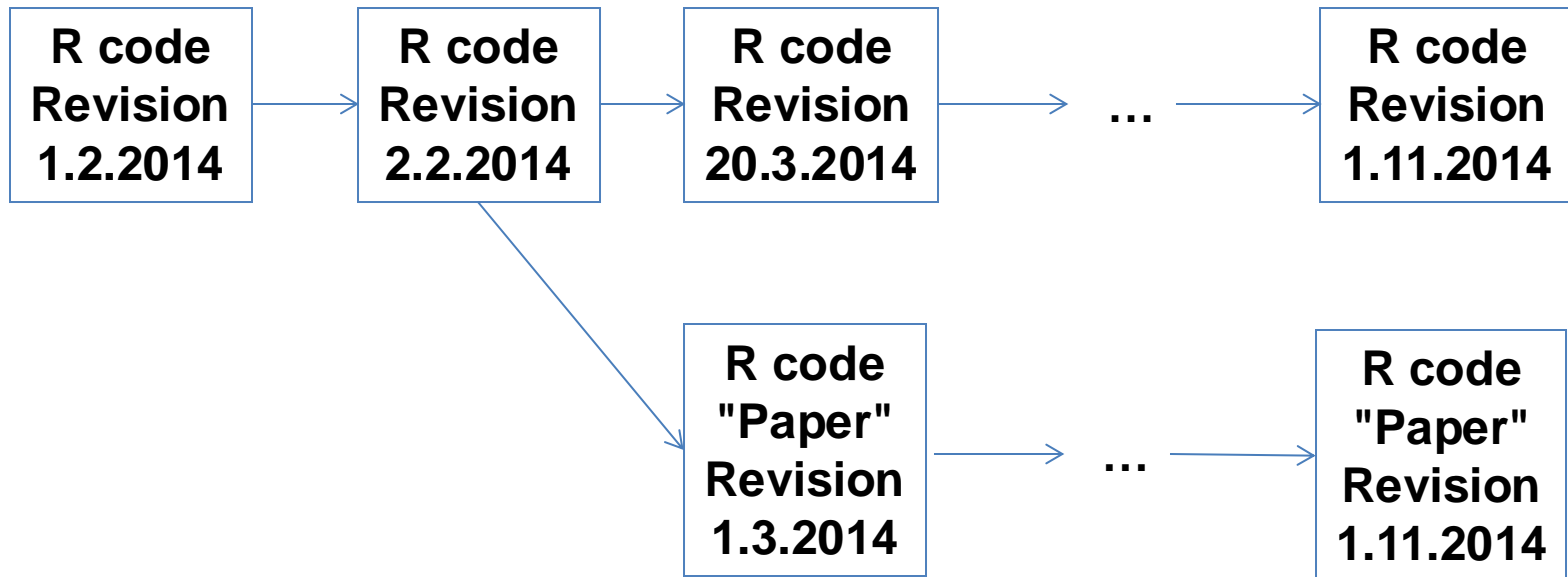


FINAL_rev.22.comments49.
corrections.10.##\$%WHYDID
ICOMETOGRADSCHOOL?????.doc

JORGE CHAM © 2012

WWW.PHDCOMICS.COM

If needed, use a tool like **Subversion** or **GIT** or subversion to manage different versions/ revisions/branches of your files.



Dynamic documents with knitr

- Based on the idea of **literate programming**
- Combine program code and explanation/documentation in same document (Donald Knuth, 1984)
- Documents in which the information is always up-to-date
- Writing report step by step while processing the data, in the same file
- Integrate your results in a report: *write the R code directly with the text, and later integrate the results directly into the text.*

Allows you to integrate your results in a report.
Write the R code directly with the text, and
later integrate the results directly into the text.

[Home](#)[Objects](#)[Options](#)[Hooks](#)[Patterns](#)[Demos](#)

knitr

Elegant, flexible and fast dynamic report generation with R



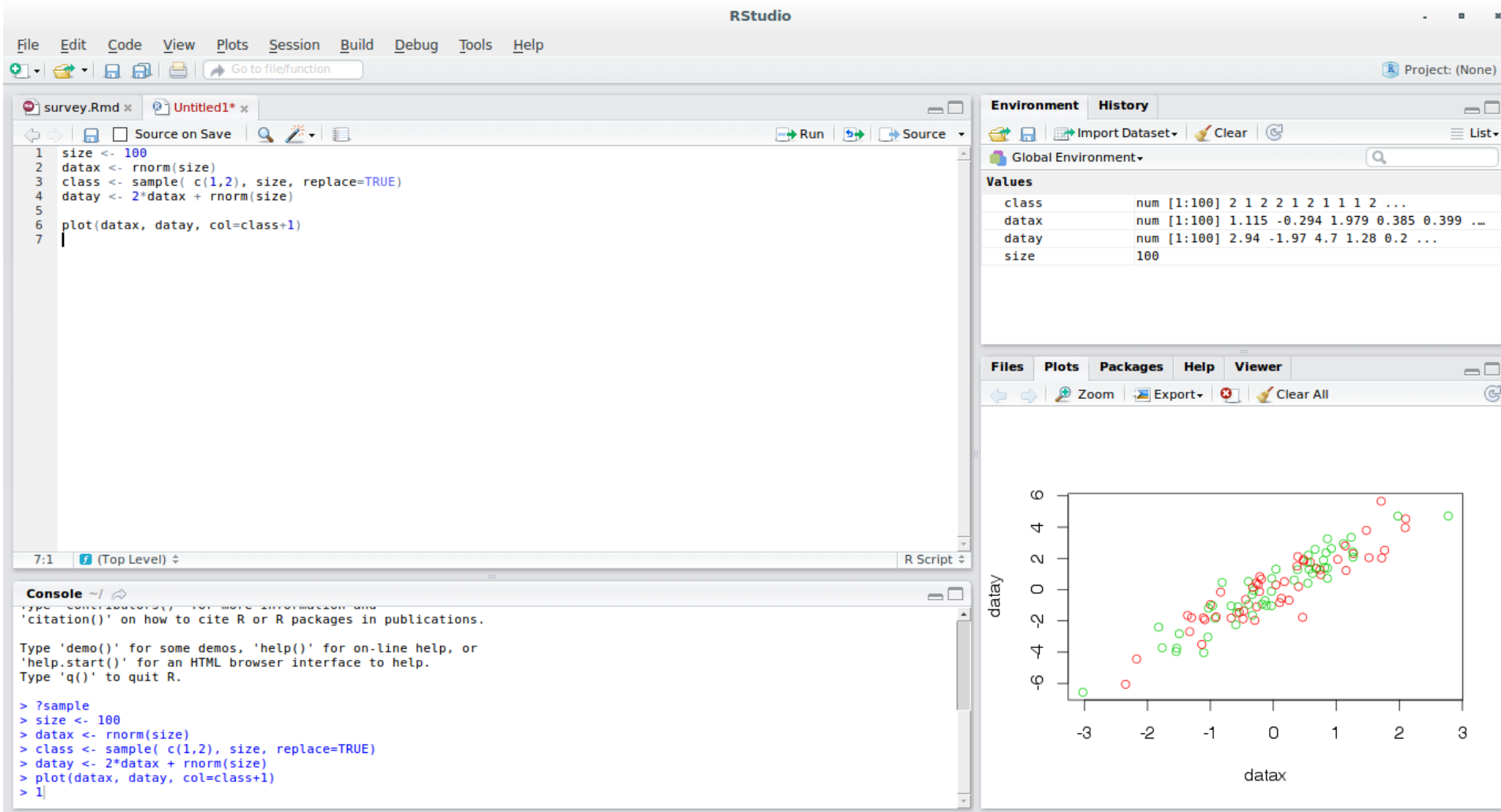
Overview

The knitr package was designed to be a transparent engine for dynamic report generation with R, solve some long-standing problems in Sweave, and combine features in other add-on packages into one package (`knitr` \approx `Sweave` + `cacheSweave` + `pgfSweave` + `weaver` + `animation::saveLatex` + `R2HTML::RweaveHTML` + `highlight::HighlightWeaveLatex` + `0.2 * brew` + `0.1 * SweaveListingUtils` + more).

<http://yihui.name/knitr/>

What we need to use knitr

- R
- `knitr` R package
- Editor (preferably with some support for R)
configured to provide support for **knitr**
see <http://yihui.name/knitr/demo/editors/>
- TeX Live (optional)
- pandoc (required for this class)
- learn from demos and examples:
 - <http://yihui.name/knitr/>
 - <http://rpubs.com>



How can we write the report ?

- Write .Rnw files, and generate PDF reports using LaTeX
- keep general structure of standard LATEX document:

```
\documentclass{...}  
\usepackage{...}  
\begin{document}  
...  
\end{document}
```
- Use the same LATEX packages/configurations as usual
- Add R chunks in the LaTeX code

If LaTeX is too scary, consider:

- **LYX:** <http://www.lyx.org/>
- **markdown:**
 - http://www.rstudio.com/ide/docs/authoring/using_markdown
 - <https://github.com/adam-p/markdown-here/wiki/>

Markdown is a simple plain text format that allows you to specify the layout of a document, and which can easily be converted to different formats afterwards.

R Markdown combines the core syntax of markdown (easy-to-write plain text format) with embedded R code chunks that are run so their output can be included in the final document.

R Markdown v2 (<http://rmarkdown.rstudio.com/>)

The screenshot displays the RStudio interface with an R Markdown document open. The left pane shows the source code, and the right pane shows the rendered HTML output.

Source Code (rmarkdown.Rmd):

```
1 ---
2 title: "RMarkdown example"
3 author: "Frédéric Schütz"
4 date: "19/11/2015"
5 output: html_document
6 ---
7
8 This is an R Markdown document. Markdown is a simple
9 formatting syntax for authoring HTML, PDF, and MS Word
10 documents. For more details on using R Markdown see
11 <http://rmarkdown.rstudio.com>.
12
13 # Introduction
14
15 Markdown allows you to create sections and subsections,
16 to emphasize text by formatting it in italic or bold.
17
18 ## Lists
19
20 You can also easily create lists:
21
22 * Item 1
23 * Item 2
24   + Item 2a
```

Rendered HTML Output:

RMarkdown example

Frédéric Schütz

19/11/2015

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

Introduction

Markdown allows you to create sections and subsections, to emphasize text by formatting it in *italic* or **bold**.

Lists

You can also easily create lists:

- Item 1
- Item 2
 - Item 2a

1. Ordered list
2. Another item
 - A subitem

or even tables:

First Header	Second Header
Content Cell	Content Cell
Content Cell	Content Cell

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

rmarkdown.Rmd

Knit HTML Run Chunks

```
33
34
35 # Integrating R code
36
37 When you click the Knit button a document will be
38 generated that includes both content as well as the
39 output of any embedded R code chunks within the
40 document. You can embed an R code chunk like this:
41
42 ```{r}
43 summary(cars)
44 ```
45
46 You can also add R code directly in the text; for
47 example, the dataset cars contains `r nrow(cars)` rows.
48
49 You can also embed plots, for example:
50
51 ```{r, echo=FALSE}
52 plot(cars)
53 ```
54
55 Note that the `echo = FALSE` parameter was added to the
56 code chunk to prevent printing of the R code that
57 generated the plot.
```

43:41 (Top Level) R Markdown

Console

~/Documents/sib/reproducible_research/rmarkdown.html

rmarkdown.html Open in Browser Publish Find

Integrating R code

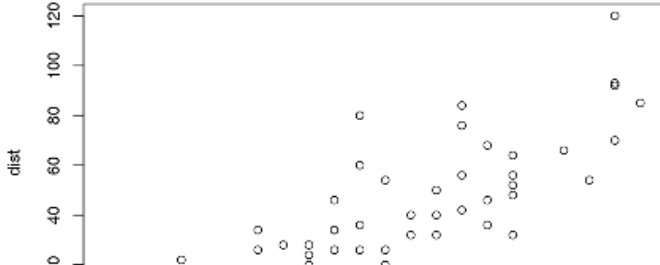
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

##	speed	dist
## Min.	: 4.0	Min. : 2.00
## 1st Qu.:	12.0	1st Qu.: 26.00
## Median :	15.0	Median : 36.00
## Mean :	15.4	Mean : 42.98
## 3rd Qu.:	19.0	3rd Qu.: 56.00
## Max. :	25.0	Max. : 120.00

You can also add R code directly in the text; for example, the dataset cars contains 50 rows.

You can also embed plots, for example:



A scatter plot showing the relationship between speed (x-axis) and distance (y-axis) for the 'cars' dataset. The x-axis ranges from 0 to 25, and the y-axis ranges from 0 to 120. The plot shows a positive correlation, with points scattered across the range, indicating that as speed increases, the distance traveled also tends to increase.

R Markdown example

```
title: "Untitled"
author: "Frédéric Schütz"
date: "23/01/2015"
output: html_document
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

R Markdown example (continued)

```
```{r}  
summary(cars)
```
```

You can also embed plots, for example:

```
```{r, echo=FALSE}  
plot(cars)
```
```

Note that the ``echo = FALSE`` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

- **Emphasis:** `*italic*` `**bold**`
 `_italic_` `__bold__`

- **Headers**

Header 1

Header 2

Header 3

- **Unordered List:**

* Item 1

* Item 2

 + Item 2a

 + Item 2b

- **Ordered list:**

1. Item 1

2. Item 2

3. Item 3

 + Item 3a

 + Item 3b

RStudio

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

survey.Rmd x Untitled1* x

Source on Save Run Source

```
1 size <- 100
2 datax <- rnorm(size)
3 class <- sample( c(1,2), size, replace=TRUE)
4 datay <- 2*datax + rnorm(size)
5
6 plot(datax, datay, col=class+1)
7
```

Environment History

Global Environment

Values

| | | |
|-------|-------------|------------------------------------|
| class | num [1:100] | 2 1 2 2 1 2 1 1 1 2 ... |
| datax | num [1:100] | 1.115 -0.294 1.979 0.385 0.399 ... |
| datay | num [1:100] | 2.94 -1.97 4.7 1.28 0.2 ... |
| size | | 100 |

Files Plots Packages Help Viewer

Zoom Export Clear All

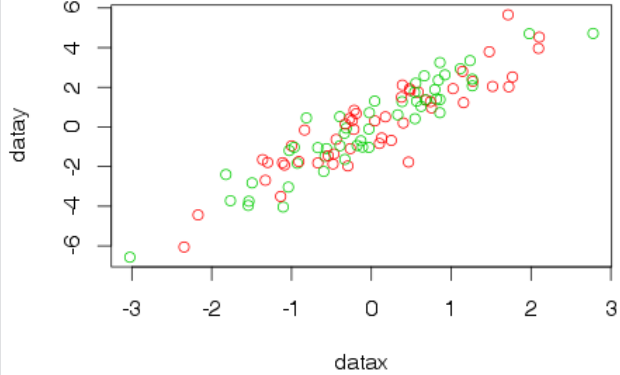
7:1 (Top Level) R Script

Console

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

```
> ?sample
> size <- 100
> datax <- rnorm(size)
> class <- sample( c(1,2), size, replace=TRUE)
> datay <- 2*datax + rnorm(size)
> plot(datax, datay, col=class+1)
> 1
```

datay



datax

The scatter plot shows 'datay' on the y-axis (ranging from -6 to 6) and 'datax' on the x-axis (ranging from -3 to 3). Data points are colored based on the 'class' variable: class 1 (green) and class 2 (red). The points are scattered across the plot area, with a slight positive correlation visible.

<http://www.rstudio.com/products/rstudio/download/>

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

report.Rmd*

Knit Word

- Knit HTML
- Knit PDF
- Knit Word

```
1
2
3
4
5
6
7
8 |---
9 title: "Statistics analysis for paper Ludigsa et al."
10 author: "Frédéric Schütz"
11 date: "4 November 2014"
12 output: word_document
13 ---
14
15 # Figure 1a (new)
16
17 The p-values are obtained using a two-sample (two-sided) t-test. Adjustment for
18 multiple testing is done using a Bonferroni correction over 12 samples.
19
20 ```{r echo=FALSE}
21
22 dotable <- function( datafile ) {
23
24   data <- read.table(datafile, fill=TRUE, header=FALSE,
25                      sep="\t", as.is=TRUE)
26
27   results <- NULL
28
29   for (group in unique(data$V1)) {
30
31     # Create section here !
32
33     for (group2 in unique(data$V2)) {
34       for (genotype in c("Nlrc5", "Rfx5")){
35
36         thisdata <- data[ order( genotype, data$V3 ), 1
```

8:1 (Top Level) R Markdown

Environment History

To Console To Source

date "4 November 2014"

Files Plots Packages Help Viewer

Zoom Export Clear All

Console ~/sibtmp/guerda_greta/

citation() on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

- R code placed in *chunks* will be evaluated and printed

```
` `` {r}  
summary(cars$dist)  
summary(cars$speed)  
` ``
```

- **Inline R Code**

There were `r nrow(cars)` cars studied

- **Links: use a plain http address or add a link to a phrase:**

```
http://example.com  
[linked phrase] (http://example.com)
```

- **Images on the web or local files in the same directory:**

```
![alt text] (http://example.com/logo.png)  
![alt text] (figures/img.png)
```

Why use knitr ?

- all-in-one: analysis, documenting, formatting, reporting
- no annoying and error-prone copy-pasting
- modifying input data or code: changes are directly reflected in report
- easy to display underlying code in report when needed
- split code in chunks, but can still access all previously defined
- variables (single R session)
- flexible: code externalization, child documents, caching,...

R notebooks

Exercises

- Using Rstudio, start a new .Rmd (R Markdown file).
- Look at the template that was provided, change the R code
- Create an HTML, a Word and a PDF file from this Markdown code
- Make sure you know how to do at least the following: sections, lists, create a table from R, insert a graphic from R.
- Note: you may need to install a TeX distribution to generate PDF; you can also generate a Word or Excel document, and print/convert them to PDF if required
- Make sure to include information about the current R session (R version, packages loaded) in the final document
- Adapt an R script of your choice (ideally one you would use in your work) in a Markdown report

